

## GERMAN CREDIT RISKS CLASSIFICATION USING SUPPORT VECTOR MACHINES

Burak TEZCAN<sup>1</sup>, Sakir TASDEMİR<sup>1, +</sup>

<sup>1</sup>Selçuk University, Department of Computer Engineering, Konya Turkey

btezcan@selcuk.edu.tr, stasdemir@selcuk.edu.tr

### Abstract

Support Vector Machines (SVM) is one of the most popular classification algorithms. SVM penalty parameter and the kernel parameters have high impact over the classification performance and the complexity of the algorithm. So, this brings the problem of choosing the suitable values for SVM parameters. This problem can be solved using meta-heuristic optimization algorithms. Salp Swarm Algorithm (SSA) and Crow Search Algorithm (CSA) are new meta-heuristic algorithms. SSA is a swarm algorithm that is inspired from a mechanism salps forming in deep ocean called salp chain. CSA algorithm is inspired by the intelligent behavior of crows. In this paper, SVM parameter optimization is done using SSA and CSA. German Credit dataset from the UCI data repository is used for the experiments. All experiments results are gathered from a 10-fold cross validation block. Evaluation criteria determined as accuracy, sensitivity, specificity and AUC. SSA and CSA gave accuracy results of  $0.72 \pm 4.62$  and  $0.71 \pm 3.53$  respectively. Also, ROC curves and box plots of the algorithms are given. CSA algorithm draws better graphs.

**Keywords:** Support Vector Machines, Optimization, Parameter, Metaheuristics

### 1. Introduction

Support Vector Machines (SVM) is a learning methodology based on Structural Risk Minimization (SRM). SVMs can give good results on non-linear problems, but SVM performance highly depends on suitable parameters. Parameters directly affects the model performance. Therefore, Particle Swarm Optimization (PSO), Genetic Algorithm (GA) and

---

<sup>+</sup> This paper has been presented at the ICENTE'18 (International Conference on Engineering Technologies) held in Konya (Turkey), October 26-28, 2018.

Grid Search (GS) have been used numerously in parameter optimization of SVMs [1-3]. However, GS algorithm is time consuming and PSO and GA algorithms often stuck in local optimums. Therefore, SVM parameter optimization needs new methods.

Nowadays, optimization is used in many fields. Conventional methods are used for simple optimization problems, but computers are used for solving high level optimization problems. Many algorithms were developed for solving optimization problems. Each algorithm has advantages and disadvantages for any problem. Many different test problems are used in literature for testing performances of these algorithms. Because of the high usage of these problems, they became benchmarks. However, in real life situations, performances can be different from the ones achieved over benchmark problems.

Optimization is finding the best solution over all solutions in given conditions. Any problem with constraints involving unknown parameter values can be called an optimization problem [4].

Sometimes, creatures that doesn't show any value by themselves can show great intelligence when they group up. Individuals belonging to a group make use of the behavior of the best individual or the all other individuals or their own experiences and use these as a tool to solve future problems. For example, an animal in a flock can react to a danger and this reaction moves in the flock to ensure all animals behave the same way against that danger. By observing these behaviors of animals, swarm intelligence algorithms are developed [5].

Salp Swarm Algorithm (SSA) is a recently developed meta-heuristic algorithm [6]. SSA has the advantages of few parameters and strong global search. In this study, SVM parameter optimization has done using SSA and CSA [7]. German Credit dataset from UCI repository used for the experiments.

Organization of the paper as follows: SVMs are defined, Swarm intelligence algorithms introduced, experiments and conclusion.

## **2. Support Vector Machines**

Consider  $X = \{x_1, x_2, \dots, x_m\}$  as a training set.  $y_i = \{-1, +1\}$  corresponds to class values. Function  $f : X \rightarrow \{\pm 1\}$  must be solved to find classes.

Structural Risk Minimization based SVMs try to find most suitable hyperplane between classes. While doing this, SVMs try to establish balance between exploitation and exploration. A class of hyperplanes are defined in search space  $H$  in Eq. 1 where  $w, x \in H, b \in R$ .

$$\langle w, x \rangle + b = 0 \quad (1)$$

Eq. 2 represents decision function.

$$f(x) = \text{sgn}(\langle w, x \rangle + b) \quad (2)$$

Vapnik proposed a method for finding the optimal hyperplane so that error rate in training set can be minimized. Eq. 3 should be solved to find optimal hyperplane. Eq 3. has the constraints in Eq. 4.

$$\text{minimize } \tau(w) = \frac{1}{2} \|w\|^2 \quad (3)$$

$$y_i(\langle w, x_i \rangle + b) \geq 1 \quad \forall i \in \{1, \dots, m\} \quad (4)$$

Using the constraints in Eq. 4, for every  $y_i = +1$ ,  $f(x_i)$  becomes +1 and  $y_i = -1$ ,  $f(x_i)$  becomes -1. Detailed information about these formulas can be found in Scholkopf and Smola's work [8].

Upper method can only be applied to linearly separable spaces. Boser ve ark. [9] proposed a kernel-based approach for the non-linear spaces where maximal hyperplane is needed. It suggests changing scalar products in Eq. 4 with a non-linear kernel function (Eq. 5).

$$y_i(K(w, x_i) + b) \geq 1 - \varepsilon_i, \quad \forall i \in \{1, \dots, m\} \quad (5)$$

Most popular kernel functions are given below:

- Linear:  $K(x_i, x_j) = (x_i \cdot x_j)$
- Polynomial:  $K(x_i, x_j) = (\gamma x_i \cdot x_j + c)^d$
- Radial Basis Function (RBF):

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$$

Here,  $x_i, x_j$  represents examples,  $d$  represents polynomial degree ve  $\gamma$  represents gauss value.

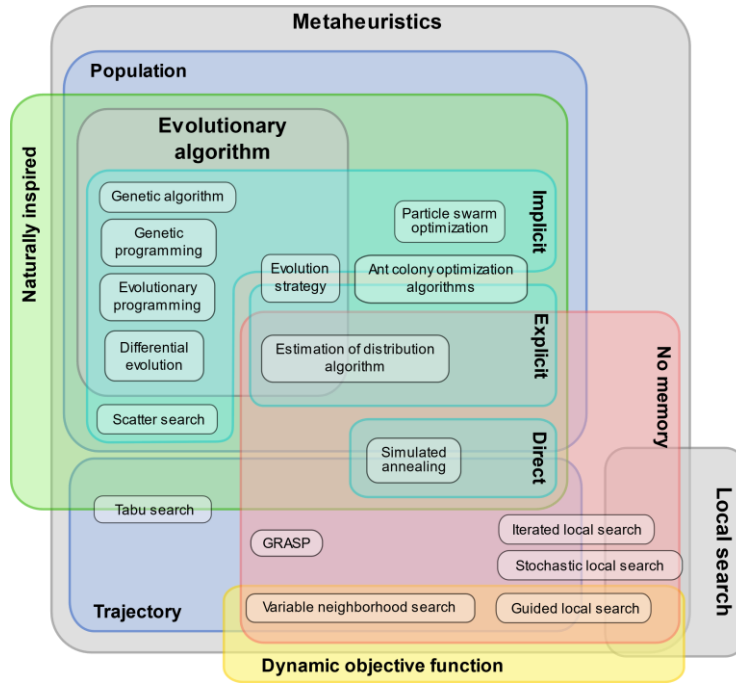


Figure 1. Euler diagram of the different classifications of metaheuristics [10]

### 3. Optimization and Swarm

Real time optimization problems are complicated and hard to solve. Generally, algorithms used in solving hard optimization problems have high computational burden and specifically design for a certain problem. Using these algorithms for different optimization algorithms is almost impossible. Therefore, heuristic algorithms are designed. Heuristic algorithms do not guarantee the best solution but works faster.

Heuristic algorithms evaluate the search space and finds a solution very close to the best. But they do not guarantee finding the best solution. When these types of algorithms are developed, they use some information about the problem they are developed for, so they have some problem specific features and called heuristic algorithms. A\* search, hill climbing algorithm and best first search are a few of the heuristic algorithms.

Metaheuristic algorithms are not problem specific. 'Meta' mean higher level in Greek. Metaheuristic algorithms can be denoted as higher-level heuristic algorithms. Metaheuristics are generally nature inspired and can be used for many different problems. Metaheuristics act like a black-box because they do not need specific information about the

optimization problem. Genetic Algorithm (GA), Ant Colony Optimization (ACO), Artificial Bee Colony (ABC) and Particle Swarm Optimization (PSO) are a few of the metaheuristic algorithms. Figure 1 shows the classification of metaheuristics.

Characteristic of metaheuristics can be given [11]:

- Metaheuristics are strategies that “guide” the search process.
- The goal is to efficiently explore the search space in order to find (near-) optimal solutions.
- Techniques which constitute meta-heuristic algorithms range from simple local search procedures to complex learning processes.
- Metaheuristic algorithms are approximate and usually non-deterministic.
- They may incorporate mechanisms to avoid getting trapped in confined areas of the search space.
- The basic concepts of metaheuristics permit an abstract level description.
- Metaheuristics are not problem-specific.
- Metaheuristics may make use of domain-specific knowledge in the form of heuristics that are controlled by the upper level strategy.
- Today’s more advanced metaheuristics use search experience (embodied in some form of memory) to guide the search.

Swarm intelligence algorithms are flexible and solid method that are developed inspired by animals’ swarm behaviors. ACO and PSO are two of the most used swarm intelligence algorithms. ACO algorithm mostly used in solutions of combinatorial optimization problems and PSO algorithm mostly used in continuous optimization algorithms. For example, routing problems (traveling salesman, vehicle routing etc.), assignment problems (graph coloring etc.), scheduling problems (open-shop scheduling etc.) can be solved using ACO and problems that needs function optimization in many different engineering fields can be solved using PSO.

Swarm can be defined as discrete individuals influencing each other. Individuals can be a human or an ant. In swarms, N individual work together to achieve a purpose. This easily observable “collective intelligence” arises from repetitive behaviors of individuals.

#### 4. Experiments

In this study, SVM parameter optimization over German Credit data has done using SSA and CSA algorithms. The two algorithms compared with each other and the literature. RBF (Radial Basis Function) kernel function were used in SVM. Two parameters of SVM were optimized. These are balancing parameter between error rate and generalization called  $C$  and RBF kernel parameter  $\gamma$ . Every population member in the optimization algorithms are defined as a combination of  $C$  and  $\gamma$ . An SVM block were used as fitness functions of the optimization algorithms. Parameters that provide best SVM accuracy were stored in each iteration. Best parameters were given when the end criterion. These best parameters were fed into a 10-fold cross validation SVM block and results were gained.

German dataset classifies people described by a set of attributes as good or bad credit risks. It includes 1000 instances and 24 attributes. The original dataset, in the form provided by Prof. Hofmann, contains categorical/symbolic attributes. For algorithms that need numerical attributes, Strathclyde University produced a numerical file. This file has been edited and several indicator variables added to make it suitable for algorithms which cannot cope with categorical variables. Several attributes that are ordered categorical (such as attribute 17) have been coded as integer.

Table 1. SSA and CSA performance results

Algorithm	Performance Criteria	Results
SSA	Accuracy	0.72±4.62
	Sensitivity	0.41±0.03
	Specificity	0.97±0.02
	AUC	0.63±0.04
CSA	Accuracy	0.71±3.53
	Sensitivity	0.43±0.09
	Specificity	0.82±0.02
	AUC	0.70±0.05

In Table 1, accuracy, sensitivity, specificity and AUC values of SSA and CSA algorithms are given. SSA algorithm gave 72% accuracy rate and it's better than the CSA's 71% accuracy rate. Both algorithms gave low sensitivity values.

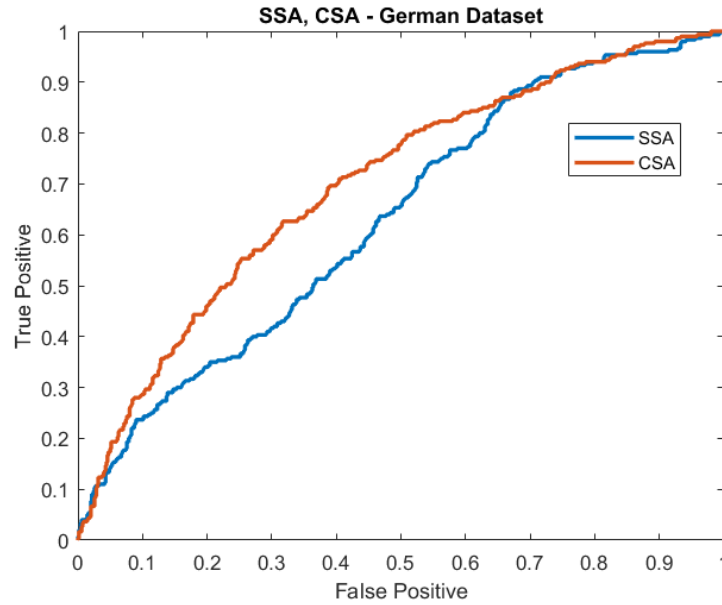


Figure 2. ROC curves of SSA and CSA

In Figure 2, ROC curves for SSA and CSA over 10-fold cross validation can be seen. CSA algorithm show a better ROC curve than SSA algorithm. CSA curve is in the area accepted as “normal” but SSA curve is close to the “bad” area. In Figure 3, boxplots of SSA and CSA can be seen. SSA boxplot draws a narrower box. CSA boxplot is higher than SSA boxplot. It can't be concluded for sure, but it can be said that there is a possibility CSA has better distribution than SSA.

## 5. Conclusion

In this study, German credit risks classification has done using Support Vector Machines. SSA and CSA algorithms used for the parameter optimization of SVMs. RBF kernel function used in SVM experiments.  $C$  and  $\gamma$  parameters were optimized. 10-fold cross validation average accuracy values were used as fitness functions of optimization algorithms. Both algorithms achieved similar results, but CSA algorithm draws better ROC curve and

boxplot. Experiments show that both algorithms can compete for SVM parameter optimization.

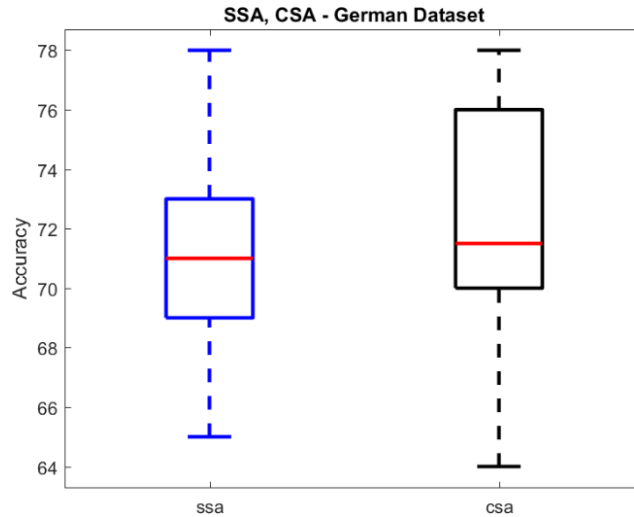


Figure 3. Boxplots of SSA and CSA

In future studies, we will consider different SVM kernel functions. Even though RBF is the most used kernel function, it can't always be better than other kernels like sigmoid, polynomial etc.

## References

- [1] Cheng, J., et al., *Temperature drift modeling and compensation of RLG based on PSO tuning SVM*. Measurement, 2014. **55**: p. 246-254.
- [2] Gencoglu, M.T. and M. Uyar, *Prediction of flashover voltage of insulators using least squares support vector machines*. Expert Systems with Applications, 2009. **36**(7): p. 10789-10798.
- [3] Li, X.Z. and J.M. Kong, *Application of GA-SVM method with parameter optimization for landslide development prediction*. Nat. Hazards Earth Syst. Sci., 2014. **14**(3): p. 525-533.
- [4] G. Murty, K., *Optimization for decision making. Linear and quadratic models*. 2009.
- [5] Akyol, S. and B. Alataş, *Güncel sürü zekası optimizasyon algoritmaları*. Vol. 1. 2012.
- [6] Mirjalili, S., et al., *Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems*. Advances in Engineering Software, 2017. **114**: p. 163-191.



- [7] Askarzadeh, A., *A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm*. Computers & Structures, 2016. **169**: p. 1-12.
- [8] Scholkopf, B. and A.J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. 2001: MIT press.
- [9] Boser, B.E., I.M. Guyon, and V.N. Vapnik, *A training algorithm for optimal margin classifiers*, in *Proceedings of the fifth annual workshop on Computational learning theory*. 1992, ACM: Pittsburgh, Pennsylvania, USA. p. 144-152.
- [10] *Metaheuristics classification.svg*, M. classification.svg, Editor. 2018, Wikimedia Commons, the free media repository.
- [11] Blum, C. and A. Roli, *Metaheuristics in combinatorial optimization: Overview and conceptual comparison*. ACM Comput. Surv., 2003. **35**(3): p. 268-308.